Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Operations Research Letters 32 (2004) 36–40
Operations Research Letters
www.elsevier.com/locate/dsw

On the complexity of the robust spanning tree problem with interval data

Ionuţ D. Aron[1], Pascal Van Hentenryck*

Department of Computer Science, Brown University, Box 1910, Providence, RI 02912, USA

Received 11 December 2002; accepted 4 April 2003

Abstract

This paper studies the complexity of the robust spanning tree problem with interval data (RSTID). It shows that the problem is NP-complete, settling the conjecture of Kouvelis and Yu, and that it remains so for complete graphs or when the intervals are all $[0, 1]$. These results indicate that the difficulty of RSTID stems from both the graph topology and the structure of the cost intervals, suggesting new directions for search algorithms.

Keywords: Robust spanning tree; Robust optimization; Uncertainty; Interval data

1. Introduction

Many practical applications, particularly in telecommunication networks, can be formulated as *robust optimization* problems. The goal in robust optimization is to find a solution that hedges against the worst possible scenario.

Several *robustness criteria* may be used when choosing among solutions, depending on the goal and specifics of the problem. Kouvelis and Yu [9] discuss three important such criteria, namely *absolute robustness*, *robust deviation* and *relative robustness*. In many cases, under these criteria, the robust equivalent of a polynomially solvable problem becomes NP-hard. A weaker criterion for robustness, which allows controlling the degree of conservatism of the solution, was proposed by Bertsimas and Sim [3]. While the three criteria in [9] aim at finding a solution whose worst case *value* is optimal, the one in [3] targets a solution with an optimal worst case *variance* (in other words, the cost of the solution is of no importance, only its stability matters). This approach enables robust formulations of polynomially solvable (α-approximable) problems to remain polynomially solvable (α-approximable).

Our study is focused on the robust spanning tree problem in graphs where the edge costs are given by intervals (RSTID), under the *robust deviation* framework. The goal of RSTID (under this framework) is to find a spanning tree which minimizes the maximum deviation of its cost from the costs of the minimum spanning trees obtained for all possible realizations of the edge costs within the given intervals. The problem has attracted considerable attention

* Corresponding author. Tel.: +1-401-8637634; fax: +1-401-8637657.

E-mail addresses: ia@cs.brown.edu (I.D. Aron), pvh@cs.brown.edu (P. Van Hentenryck).

[1] Currently on leave at GSIA, Carnegie Mellon University.

0167-6377/03/$ - see front matter
doi:10.1016/S0167-6377(03)00058-0

recently [2,9,10,15,20], particularly because of its importance in communication networks. For instance, as suggested in [9], it can be used to model uncertainty of routing delays on the edges of a network. A robust spanning tree is desirable in this case, as it provides a backbone on which routing (multicasting) delays are as close as possible to the minimal delays under any traffic load.

It is important to point out that the choice of a robustness criterion affects not only what the solution looks like, but also the complexity of finding such a solution. In the case of RSTID, a polynomial time algorithm exists under the *absolute robustness* framework [20], but none was found for the *robust deviation* criterion. In fact, Kouvelis and Yu [9] conjectured that no such algorithm exists and essentially set the tone for all subsequent attempts to solve the problem, which include a mixed integer programming (MIP) formulation [20], a constraint satisfaction approach [2] and a branch and bound approach [15].

This paper closes this conjecture by showing that RSTID is at least as hard as the *central tree* problem, introduced by Deo [5] in the late sixties and proved NP-complete three decades later, in [4]. Further, it proves that RSTID remains NP-complete even in the case of complete graphs, in contrast to the central tree problem, which admits efficient solutions for such graphs [11]. These results shed a new light on the algorithms that have been proposed for the RSTID problem [2,15,20]. These algorithms have focused almost exclusively on the cost structure and have ignored the topological properties of the graph. Our results show that both the cost structure and the topology of the graph contribute to the complexity of the problem and, as a consequence, they provide new directions to improve these algorithms and to evaluate them experimentally.

The rest of the paper is organized as follows. Section 2 gives a formal definition of the RSTID problem. Section 3 summarizes prior work on central trees, which dates back to 1966. Section 4 proves that the special case of RSTID, where all the cost intervals are $[0,1]$, is NP-complete. Section 5 shows that RSTID remains NP-complete even on complete graphs and Section 6 concludes the paper.

2. The RSTID problem

The RSTID problem is defined on an undirected, connected graph $G=(V,E)$, whose edges $e \in E$ are associated with cost intervals $[\underline{c_e},\overline{c_e}]$. A *scenario* is a particular assignment of costs to edges from their corresponding intervals. The *deviation* of a spanning tree T under a given scenario s, denoted by $\Delta_T(s)$, is the difference between the cost of T and the cost of a minimum spanning tree under scenario s, where the cost of a tree is simply the sum of its edge costs:

$$\Delta_T(s) = cost(T,s) - cost(MST_s,s). \tag{1}$$

A scenario s_T^* under which this difference is maximized is called a *worst case scenario* for the spanning tree T and the corresponding deviation is the *robust deviation* of T, which is denoted by Δ_T^*. In symbols,

$$s_T^* \in \arg\max_s \Delta_T(s) \tag{2}$$

and

$$\Delta_T^* = \Delta_T(s_T^*). \tag{3}$$

The following result gives a characterization of a worst case scenario for a spanning tree T:

Lemma 2.1 (Yaman et al. [20]). *The robust deviation of a spanning tree T is achieved when edges in T are at the highest cost and edges in the complement of T in G are at the lowest cost (i.e. $c_e=\overline{c_e}, \forall e \in T$ and $c_e=\underline{c_e}, \forall e \in E \setminus T$).*

A *robust spanning tree* is a spanning tree whose robust deviation is minimum. The *robust spanning tree problem with interval data* is defined as the following decision problem: *Given an undirected and connected graph G with interval edge costs and a value $D \in \mathbf{R}_+$, is there a spanning tree T of G whose robust deviation $\Delta_T^* \leqslant D$?*

3. Central trees

The *distance* between two spanning trees of a graph is defined as the number of edges present in one tree but not in the other [19]:

$$d(T_1,T_2) = |T_1 - T_2| = |T_2 - T_1|. \tag{4}$$

The authors of [4] and [5] use slightly different definitions for the notion of distance between trees, but it is easy to see that they are equivalent, since the notion refers only to complete (i.e., spanning) trees. Recall that the cospanning tree of a spanning tree T is the edge complement of T in G. Also, the *rank* $\rho(G)$ of a graph G with n vertices and k connected components is $n-k$.

A *central tree* [5] of a graph G is a tree T_0 such that the rank r of its cospanning tree $\overline{T_0}$ is minimum, i.e.,

$$r = \rho(\overline{T_0}) \leqslant \rho(\bar{T}), \quad \forall T \in G. \tag{5}$$

Deo [5] pointed out that, if r is the rank of the cospanning tree of T, then there is no tree in G at a distance greater than r from T and there is at least one tree in G at distance exactly r from T. A direct consequence of this is the following characterization (and hence the name) of central trees:

Lemma 3.1 (Deo [5]). *A spanning tree T_0 is a central tree of G if and only if the largest distance from T_0 to any other tree in G is minimum, i.e.*,

$$\max_i d(T_0, T_i) \leqslant \max_i d(T, T_i), \quad \forall T \in G. \tag{6}$$

Central trees were studied intensively in the literature [12,16–18], primarily due to their importance in circuit analysis. Deo's initial study was motivated by the intuition that a central tree would provide a better starting point for tree generation methods, such as those proposed in [6,14]. Early attempts to solve the central tree problem resulted in a polynomial time algorithm proposed in [1], but a gap in this algorithm was uncovered in [7]. The problem remained open for almost three decades, in spite of some interesting similarities with problems for which efficient algorithms exist. The *maximally distant* trees problem, for instance, which asks for a pair of spanning trees (T_1, T_2) such that $d(T_1, T_2) \geqslant d(T_i, T_j)$, $\forall T_i, T_j \in G$, can be solved in polynomial time [8]. Also, as pointed out in [13], the distances between tree pairs of a graph G are in a one-to-one correspondence with the distances between the corresponding vertex pairs in the *tree-graph* $\mathscr{T}_G$ (recall that the vertices of a tree-graph $\mathscr{T}_G$ correspond to spanning trees in G, and two vertices are connected if the original spanning trees are at distance 1 of each other in G). Thus, finding a central tree in G is equivalent to finding a *central vertex* in $\mathscr{T}_G$. However, while the central vertex problem is known to have a polynomial time algorithm (in the number of vertices), such an algorithm cannot be used to efficiently find a central tree, since the number of vertices in $\mathscr{T}_G$ can be exponential. The search for an efficient algorithm for the central tree problem concluded with the following result, due to Bezrukov et al. [4]:

Theorem 3.2 (Bezrukov et al. [4]). *The central tree problem is NP-complete.*

4. The zero-one RSTID

Consider the case of RSTID in which all edges of G take their costs from the same interval, $[0, 1]$. We refer to this problem as the *zero-one* robust spanning tree problem with interval data (ZO-RSTID) associated with G.

The next theorem is the main result of the paper. It shows that robust spanning trees of ZO-RSTID are central trees of the underlying graph and vice-versa.

Theorem 4.1. *A spanning tree T is a robust spanning tree of ZO-RSTID if and only if it is a central tree of G.*

Proof. By definition, a robust spanning tree is a tree with a minimum robust deviation, so the set of optimal solutions to ZO-RSTID is

$$\mathscr{S}_1 = \arg\min_T \Delta_T^*. \tag{7}$$

On the other hand, by Lemma 3.1, the set of central trees of G is

$$\mathscr{S}_2 = \arg\min_T \max_{T'} d(T, T'). \tag{8}$$

Let T be a spanning tree of G. Consider a worst case scenario for T as given by Lemma 2.1, i.e., a scenario where all edges of T have cost 1 and the remaining edges have cost 0. Then a minimum spanning tree for this scenario is a spanning tree that uses the minimum number of edges of T. So $\Delta_T^* = |T| - \min_{T'} |T \cap T'| = \max_{T'} |T - T'| = \max_{T'} d(T, T')$, which means that $\mathscr{S}_1 = \mathscr{S}_2$. $\square$

Corollary 4.2. *ZO-RSTID is NP-complete.*

The NP-completeness of the robust spanning tree problem with interval data follows now directly from Corollary 4.2:

Corollary 4.3. *RSTID is NP-complete.*

5. RSTID on complete graphs

It is interesting to remark that, in the case of complete graphs, the central tree problem is known to have a polynomial time algorithm [11]. So, it is natural to ask whether such an algorithm exists for RSTID as well.

Yaman et al. [20] have shown that, in a graph G with interval edge costs, there exists a well defined set of trees which contains all the solutions of the RSTID problem. These trees are called *weak trees* and they are essentially the minimum spanning trees of G with respect to the scenario set. More precisely, a spanning tree T is weak if there exists at least one scenario under which T is a minimum spanning tree of G. The following result relates the solutions of RSTID to the weak trees of G:

Lemma 5.1 (Yaman et al. [20]). *A robust spanning tree is a weak tree.*

Edges appearing on weak trees are called *weak edges* and are characterized by the following property:

Lemma 5.2 (Yaman et al. [20]). *An edge e is weak if and only if there exists a minimum spanning tree using e when its cost is at the lowest bound and the costs of the remaining edges are at their highest bounds.*

Note that this property, together with Lemma 5.1, can be used to reduce the size (i.e., number of edges) of G before attempting to solve RSTID. The MIP approach proposed in [20] achieves considerable speedup by exploiting these properties to preprocess the graph. An enhanced algorithm for detecting non-weak edges was proposed in [2] and made it possible to use preprocessing for subproblems, which further reduced the time to find an optimal solution. However, this cannot change the complexity of the problem, as can be seen from the following result:

Theorem 5.3. *The RSTID problem remains NP-complete on complete graphs.*

Proof. Let $G=(V,E)$ be an arbitrary undirected and connected graph and let $G'=(V,E\cup E')$ be the complete graph resulting by adding to G the missing edges E'. Let $\mathscr{P}$ be the RSTID problem over G' given by the following edge cost intervals:

$$c_e \in \begin{cases} [0,1] & \forall e \in E, \\ [1+\varepsilon, 1+2\varepsilon] & \forall e \in E',\ \varepsilon > 0. \end{cases} \tag{9}$$

By construction, the intersection of any minimum spanning tree MST_s of G' with E' is empty, regardless of the scenario s. That is, all minimum spanning trees of G' (with respect to the set of scenarios) take edges only from E. Therefore, by Lemmas 5.1 and 5.2, a robust spanning tree of $\mathscr{P}$ must also take edges only from E. This means that the robust spanning trees of $\mathscr{P}$ are precisely the robust spanning trees of the ZO-RSTID problem associated with G. The result follows from Corollary 4.2. □

6. Conclusion

This paper reconsidered the RSTID problem and closed the conjecture of Kouvelis and Yu [9], by showing that RSTID is at least as hard as the central tree problem [5], and therefore NP-Complete [4]. Furthermore, it showed that RSTID remains hard on complete graphs, even though a central tree can be found in polynomial time on such graphs [11].

These results shed a new light on the algorithms that have been proposed for RSTID [2,15,20], since they have focused almost exclusively on the cost structure and have ignored the topological properties of the graph. An immediate consequence of Theorem 4.1 is that when all cost intervals are $[0,1]$, the complexity of RSTID is dictated by the structure of the underlying graph (note that it does not matter whether the cost interval is $[0,1]$ or $[a,b]$, since we can always shift and normalize it). In other words, if we have an efficient algorithm for the central tree problem on a given graph, then the same algorithm solves RSTID *when all edges take costs from the same interval*. The only case known so far for which such an algorithm exists is when the graph is complete [11]. An interesting question, therefore, is whether there are other classes of graphs on which a central tree can be found efficiently. On the other hand, Theorem 5.3 reveals that the topology of the graph is not the only reason why RSTID is hard (as Theorem 4.1 may seem to suggest). This raises the following question: what are the cases when one could quickly decide on a solution based *only* on the structure of the cost intervals? One such case, for instance, is when all intervals have an empty intersection. In this case, it is easy to see that any minimum spanning tree algorithm would provide a robust spanning tree. Finally, it would be interesting to understand the combined effect of the topology of the graph and the structure of the cost intervals. As we have seen, there is one case in which these aspects, *together*, make the problem easy and one might ask whether there are any other such combinations on which RSTID can be solved efficiently.

Acknowledgements

This work was partially supported by NSF ITR Awards DMI-0121495 and ACI-0121497. The authors would like to thank Shabbir Ahmed, Ismael de Farias, Michel Goemans, George Nemhauser, and Eli Upfal for interesting discussions on robust spanning trees, as well as one of the referees for suggesting the shorter and more elegant proof of Theorem 4.1 included in the current version.

References

[1] A. Amoia, G. Cottafava, Invariance properties of central trees, IEEE Trans. Circuit Theory CT-18 (1971) 465–467.

[2] I. Aron, P. Van Hentenryck, A constraint satisfaction approach to the robust spanning tree problem with interval data, in: Proceedings of the 18th Conference on UAI, Edmonton, Canada, August 1–4, 2002.

[3] D. Bertsimas, M. Sim, Robust discrete optimization, Working Paper, Operations Research Center, MIT, January 2002.

[4] S. Bezrukov, F. Kaderali, W. Poguntke, On central spanning trees of a graph, in: Lecture Notes in Computer Science, Vol. 1120, Springer, Berlin, 1996, pp. 53–58.

[5] N. Deo, A central tree, IEEE Trans. Circuit Theory CT-13 (1966) 439–440, correspondence.

[6] S.L. Hakimi, On trees of a graph and their generation, J. Franklin Inst. 272 (1961) 347–359.

[7] F. Kaderali, A counterexample to the algorithm of Amoia and Cottafava for finding central trees, Technical Report FB 19, TH Darmstadt, June 1973, Preprint.

[8] G. Kishi, Y. Kajitani, Maximally distant trees and principal partition of a linear graph, IEEE Trans. Circuit Theory CT-16 (1969) 323–330.

[9] P. Kouvelis, G. Yu, Robust Discrete Optimization and its Applications, Kluwer Academic, Dordrecht, 1997.

[10] G. Kozina, V. Perepelista, Interval spanning tree problem: solvability and computational complexity, Interval Comput. 1 (1994) 42–50.

[11] Y.Y. Lee, Use of compound matrices in topological analysis, Master's Thesis, Kansas State University, Manhatten, KS, 1968.

[12] N. Malik, On Deo's central tree concept, IEEE Trans. Circuit Theory CT-15 (1968) 283–284, correspondence.

[13] N. Malik, Y.Y. Lee, Finding trees and signed tree pairs by the compound method, in Tenth Midwest Symposium on Circuit Theory, 1967, pp. VI-5-1-VI-5-11.

[14] W. Mayeda, S. Seshu, Generation of trees without duplications, IEEE Trans. Circuit Theory CT-12 (1965) 181–185.

[15] R. Montemanni, L. Gambardella, A. Rizzoli, A. Donati, A branch and bound algorithm for the robust spanning tree problem with interval data, Technical Report No. IDSIA-10-02, Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, 2002.

[16] S. Shinoda, T. Kawamoto, Central trees and critical sets, in: D.E. Kirk (Ed.), Proceedings of the 14th Asilomar Conference on Circuit, Systems and Computers, Vol. 108, Pacific Grove, California, 1980, pp. 183–187.

[17] S. Shinoda, T. Kawamoto, On central trees of a graph, in: Lecture Notes in Computer Science, Vol. 108, Springer, Berlin, 1981, pp. 137–151.

[18] S. Shinoda, K. Saishu, Conditions for an incidence set to be a central tree, Technical Report CAS80-6, Institute of Electrical and Communication Engineering, Japan, 1980.

[19] K. Thulasiraman, M.N.S. Swamy, Graphs: Theory And Algorithms, Wiley, New York, 1992, ISBN 0-471-51356-3.

[20] H. Yaman, O. Karasan, M. Pinar, The robust spanning tree problem with interval data, Oper. Res. Lett. 29 (2001) 31–40.